DOCUMENT RESUME

ED 228 245 SP 022 449

AUTHOR Whitla, Dean K.
TITLE Value Added and Other Related Matters.
INSTITUTION National Commission on Excellence in Education (ED), Washington, DC.
SPONS AGENCY Department of Education, Washington, DC.
PUB DATE [81]
CONTRACT NIEP820071
NOTE 34p.
PUB TYPE Viewpoints (120) -- Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Gains; Achievement Tests; *College Students; *Educational Assessment; *Educational Objectives; *Educational Trends; Evaluative Thinking; *General Education; Higher Education; Interpersonal Competence; Interpersonal Relationship; Mastery Learning; Moral Development; Secondary Education; Student Development; Writing (Composition)
IDENTIFIERS National Commission on Excellence in Education; Scholastic Aptitude Test; *Value Added

ABSTRACT

The term "value added" refers to the assessment of the amount of learning that takes place during the college years. Two experiments, Value Added I and Value Added II, attempted to measure college students' attainment of eight liberal education objectives: (1) writing ability; (2) analytical ability; (3) sensitivity to ethics, morals, and values; (4) mastery of concepts across major disciplines; (5) appreciation of self, social, and universal understanding; (6) interpersonal relationships; (7) view of life experiences; and (8) broad intellectual and aesthetic interests. Results from the Value Added I experiment demonstrated that change did occur during the college years, and results from Value Added II indicated that colleges can and do add value in very large and significant ways. Findings from these two studies are reflected upon as the author reviews past and current high school and college student trends, including Scholastic Aptitude Test scores, Achievement Test scores, sources of learning, and chosen vocations of graduates. Suggestions for improving the present state of American education are made. (CJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED228245

VALUE ADDED AND OTHER RELATED MATTERS

Dean K. Whitla

Office of Instructional Research and Evaluation

Harvard University

This publication was prepared with funding from the U.S. Department of Education, under contract #NIEP82007l. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgement in professional and technical matters. Points of view or opinions do not necessarily represent those of the U.S. Department of Education.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SP 022 449

VALUE ADDED AND OTHER RELATED MATTERS

Invited Paper for the National Commission on Excellence in Education

Dean K. Whitla
Office of Instructional Research and Evaluation
Harvard University

Much of my professional life has been spent in search of ways to improve the education of undergraduates. I've diligently searched for flaws in the system as it currently operates and for mechanisms that can improve the quality of instruction and the environment in which it takes place. Let me illustrate the power of the right mechanism by taking an example from President Eliot's work, which I've always great admired. The mechanism he used was that of the departmental model. (Harvard, incidentally was, I believe, the first American university to apply this German concept.) What Eliot did was to restructure Harvard around this new approach. It had a bracing effect on scholarship. For the first time physicists judged the work of other physicists; English scholars judged the work of other English scholars; and so on. This, probably more than any other single factor, raised the quality of American scholarship to a level which was competitive for the first time with that found in Europe. With such brilliant strokes, Eliot transformed Harvard from a parochial New England college into a national university.

In my own work, the mechanisms with which I have worked have of course not had such a profound impact as those of President Eliot; but let me cite three or four that I think are particularly important not only at Harvard, but to American education in general. The first example is that of student evaluation of courses. At the end of the semester, data from courses is routinely collected using a standard student questionnaire about the nature of that parti-

cular course and also soliciting student comments. At Harvard and at hundreds of colleges around the country, these data are now published in a very complete form with comments, quantitative results, course outlines and even copies of the final examinations. Publication of evaluations has had a very salutory effect on the quality of instruction. Students claim that the evaluations are important to them in their selection of courses. In my view, the mechanism is even more important because of the feedback it provides faculty members about their teaching. While one does not want to be totally consumer-oriented, as some of my colleagues have suggested such evaluations are, it is a rare person who can ignore such evaluations and not modify his or her behavior because of the published information. There is conclusive proof that many changes have been made in the conduct of courses because of published reviews.

As a sequel to the student evaluations, with the assistance of Derek Bok, Harvard's president, some seven years ago I established a Center for Teaching and Learning which in turn has led to significant improvements in instruction. Of the four or five programs operated by the Center, one in particular, while very labor intensive, has consistently resulted in the improvement of teaching behavior. In this project, classes are videotaped, and the tapes are played back for the instructor in the presence of a counselor. Both course evaluations and videotaping of classes are examples of mechanisms that exert a lot of leverage for improving the quality of instruction.

Let me cite another example. Every three years my Office has surveyed students living in thethirteen Harvard houses. Houses at Harvard are much more than simply living accommodations for upperclasspeople. They are education units with a Master in charge who is almost always a professor; a Senior Tutor or Dean of the House; a residential tutorial staff; a Senior Common "Room" (composed of the faculty of the House). Each House has a

dining hall, a library and classroom space. Houses are designed to serve the educational as well as the living needs of students. These tri-annual evaluations have clearly shown how each of the Houses functions and these results while available publicly only in the aggregate, are reviewed individually for each House by the President, the Deans and the House Masters. The results have led to an improvement in the quality of the house experience both educationally and socially. I would be naive to suggest that in the Houses where improvements have taken place that the Masters would give credit to the evaluation as the sole impetus for change and I would agree. What the survey does is to call to everyone's attention the problem areas which need work. Effective and thoughtful Masters have made appropriate changes and the improvements have been substantial.

Let me cite a fourth example. My Office has also evaluated departments. This is a more complex process, for unlike Houses, all departments do not see their task as essentially the same. Some serve primarily undergraduates others primarily graduate students. Their approaches differ at times simply as a reflection of size. Even so, we have been able to make some useful comparisons among the departments and help them improve in their functioning. In some departments advising was found to be particularly weak, in others it was their tutorial programs, yet in others it was a poor introductory sequence of courses for concentrators. All in all, the evaluations have been very helpful, as some would say, in identifying targets of opportunity.

Let me turn now to the first major topic of this paper--Value Added. I believe it was David Riesman who first used this term in an educational

context. Those of you who have read Alexander Astin's paper written for the Commission or his editorial in a recent Chronicle of Higher Education already have an understanding of the major concepts underlying this process. Value Added, as the term implies, is an assessment of the amount of learning that takes place during the college years. What one is seeking is an understanding of the change that takes place from freshman to senior year. Why would a place such as Harvard be interested in Value Added? For decades people have claimed that the success of Harvard graduates was not a result of the education they received but simply a measure of the quality of students admitted. Clearly, Value Added is not an index designed to make a school like Harvard look strong. But Mr. Bok expressed a deep interest in knowing exactly what the University was providing for undergraduates and how we might improve those experiences. We decided that the methodology of Value Added--the measurement of change between freshman and senior year--was appropriate.

I assembled a roster of colleagues to serve as a steeering committee. It consisted of Mr. Bok; Bill Perry, known for his work on the stages of student development; Lawrence Kohlberg for his work in moral development; David McClelland who has spent his professional life developing the concepts of need achievement and need power; David Riesman, a preeminent sociologist interested in the problems of education with whom I had the privilege many years ago of serving as a teaching fellow; and Matina Horner who in addition to being President of Radcliffe is the creator of the concept of fear of success in women.

As we began the project which was generously supported by FIPSE, we defined objectives of liberal education as eight.

I. The ability to communicate in writing with clarity and style.

II. The capacity to analyze problems by collecting relevant data and marshalling pertinent arguments.

III. A sensitivity to ethical considerations and the capacity to make discriminating moral and value choices.

IV. An ability to master new concepts and materials across the major disciplines.

V. A critical appreciation for the ways we gain an understanding of the universe, society and ourselves.

VI. A sensitivity to interpersonal relationships.

VII. The extent to which life experiences are viewed in a wide context.

VIII. A broadening of intellectual and aesthetic interests.

These objectives were chosen as central objectives of liberal education and certainly were not unique to Harvard, though the particular form of the statements came from the writings of Derek Bok and Henry Rosovsky, Dean of the Faculty of Arts and Science. As I have discussed these objectives with faculties at a variety of institutions, I've been surprised, even pleased, as to the degree to which other institutions seem to accept them as a reasonable expression of their mission, too.

Having agreed upon these objectives as central to the undergraduate experience, our next task was to determine ways of measuring these qualities. This process was an experiment entitled Value Added I.

From our studies of college alumni/ae, two of the most frequently cited long-lasting effects of college are capacity for self-expression, verbal or written (70%), and ability to think clearly (66%). While it is encouraging to know that graduates believe that college fosters the development of thinking and writing skills, we wanted to be able to measure the degree to which such developments are taking place among the current undergraduates.

It seemed impossible to identify a single essay topic which would elicit

abilities to organize and present ideas without favoring students with particular backgrounds and interests. Therefore, to measure our first objective--the ability to communicate in writing with clarity and style--we presented students with five different topics from which they could select one. All essays were coded for spelling, grammar, and organizational flaws. In addition, each essay was coded for the quality of the arguments and counterarguments presented.

While we would like to report that all seniors composed more forceful and logical essays, made fewer syntactical mistakes, and even spelled better than freshmen, we cannot. Humanities and social science seniors wrote with a finer pen than did the freshmen--for them improvement was substantial especially for the humanities people. However, natural science seniors who were very able by SAT standards, did very little better than their freshmen counterparts. Professors in the Natural Sciences who taught substantial numbers of concentrators, when informed of this fact agreed that writing was very important and were disappointed that their concentrators improved so little. They discussed strategies to improve students' skills; among them assigning a paper rather than another problem set; and a science magazine was founded so that there was recognition for good writing.

To measure the second objective--students' capacity to analyze problems by collecting relevant data and mastering pertinent arguments--we constructed the "analysis of argument" test. It consisted of an excerpt selected from a sermon by the Reverend Norman Vincent Peale in which he criticized the permissive child-rearing practices advocated by Dr. Benjamin Spock, claiming that these practices were creating moral lassitude among the young, resulting in the most undisciplined age in history. After reading the article, students were asked to argue against his position. When this task was completed they were then asked to reverse their stand and argue in Peale's behalf.

A scoring system was devised which gave credits in proportion to the quality of the argument, its analytical strength and the logic of its organization.

The results were striking; the scores of seniors were significantly higher, especially when reversing their position than were those of freshmen. Seniors were more adept at mounting effective and logical arguments supporting an alien viewpoint than were freshmen. One commentator suggested semi-humorously that such flexibility might be psychologically interpreted as a lack of committment on the part of the seniors. In fact, the test would appear to be a poor measure of committment and is more accurately interpreted as a measure of analytical ability as exemplified in understanding both sides of an argument and thus comprehending an issue more fully. It is reassuring, as McClelland has suggested, to be able to demonstrate that college does improve reasoning ability.

We used Kohlberg's Test of Moral Development to examine our third objective, to determine students' capacities to analyze ethical arguments and to make discriminating value choices. According to Kohlberg, the three basic levels of moral development are 1) the pre-conventional level, using essentially obedience and punishment modes of behavior; 2) the conventional level, based on the performance of good or correct roles; and 3) the post-conventional level, using conformity to self ideals and shared norms. Each of these basic levels has two stages. Stages one and two, at the pre-conventional level are highly egotistic. Stage three, where the conventional level begins, is symbolized by the good boy/good girl syndrome and stage four is a continuation of authority and social order. Stage five begins the search for shared norms but still operates in a contractual-legalistic framework. Stage six is the conscience or the principle orientation stage. The data Kohlberg used to determine the stages of moral development come

from situations in which students are asked to resolve moral dilemmas. These dilemmas have previously been given as oral interivews, but were transformed as part of this research into a written format which contained sufficient elaboration so it could be scored according to the Kohlberg stages.

Scoring turned up significant differences between freshmen and upperclassmen. Typically men attained higher scores than women; at Harvard however, there were no sex differences. While social science concentrators has the highest mean score, and natural science concentrators the lowest, differences by field of concentration were not significant.

Although upperclassmen had higher scores than freshmen on the Moral Judgement Test, the absolute difference was not impressive from the standpoint of stage theory. Why did seniors differ so little from freshmen? Kohlberg argues that formal operations (Piaget's concept) are necessary though not sufficient for the growth of moral judgement. This argument seems reasonable, since moral judgement develops as both a function of the ability to recognize moral issues in a hypothetical situation and the ability to reason analytically and generate principles. Therefore movement from Stage 3 to Stage 4 requires only an elaboration of role-taking to include a larger social context; the principled reasoning involved in Stage 5, however, requires, according to Kohlberg, "a clear effort to reach a personal definition of moral values--to define principles that have validity and application apart from the authority of groups or persons and apart from the individual's own identification with these groups." A possible explanation for small progress at this level suggests that it may be simply more difficult to move from Stage 4 to Stage 5 than it is to move from Stage 3 to Stage 4.

Because of the complexity of scoring the Kohlberg scales, in our more recent work we have substituted the Defining Issues test, a test derived

from Kohlberg's work by Jim Rest at the University of Minnesota which puts Kohlbergian concepts into a more easily scored format.

The ability to master new concepts and materials across the major disciplines, our fourth objective, was measured by a Learning New Material test. The aim of the test was to determine if students were more capable of learning new material as a result of college experiences. The learning task consisted of a brief essay on DNA which contained three foci: one paragraph advanced the scientific principles involved, another to social science interpretation of DNA research and a third, the humanistic meaning of such work.

The findings proved that seniors did not learn at a significantly faster rate than freshmen. Their scores were, in fact, correlated significantly with their SAT scores. This attempt to use another approach to assessing learning only corroborated the traditional finding--that learning rate does not generally increase during the college years. There was, however, an important exception: learning rate was influenced positively by the college major. Students concentrating in the natural sciences learned material in the natural sciences more rapidly than they learned material in the social science and the humanities; the same phenomenon was also true in the other areas. It appears that a facility is developed within a discipline that makes it possible for students to learn new materials within their domain more quickly than those outside it. So, not only does their choice of concentration lead to a mastery of a discipline itself, but it may well influence their lives over a longer period because of their increased ability to learn material within that field.

Robert Rosenthal developed a test to provide a measure of a subject's sensitivity to visual and auditory stimuli. It seemed an appropriate way to examine our fifth objective, evaluating students' sensitivity to inter-

personal relationships. The tests consisted of an audio tape which altered normal speech in two ways. In the first section, the frequency bandwidth was reduced so that words themselves were incomprehensible although the tonality of the speech remained. The listener had to rely on his or her intuitive understanding to ascertain what emotion was being expressed. the second portion of the tape, the natural range of the speech frequency remained, but the order of the syllables was rearranged. In order to understand the words, a high degree of analytical ability was required to take bits and pieces of the soliloquy and reassemble them into something understandable. Students were asked to identify from the tone of these snippets of incomprehensible speech whether, for example, it was the voice of someone nagging a child or saying a prayer.

Bales and Rosenthal found that a group dynamics course designed to increase interpersonal sensitivity (Harvard's Psychology and Social Relations 1200) did increase scores on this test. In another context, McClelland found that United States Information officers judged as outstanding in their work abroad had significantly higher Profile of Non-Verbal Sensitivity test scores than did officers whose work was judged average.

Women, in general, score higher than men on the Profile on Non-Verbal Sensitivity Test. Matina Horner feels that this extra sensitivity as well as an aid can also be a handicap for women. They may allow their sensitivity to people to keep them from pushing ahead for achievements that they might otherwise make. We would all benefit if men, on the other hand, were more thoughtful and sensitive to those around them. President Horner feels that a better balance should result from the college experience. The change over the college years tended to fulfill one half of President Horner's hope; the results indicate that women became less sensitive, but

men did not become more so. Rather than innumerating our procedures for evaluating all of our objectives let me turn to what meaning all of this had.

Value Added I clearly demonstrated that change occurred during the undergraduate years, but it did not provide the causes underlying these changes. Our second exploration, entitled Value Added II, was an experiment to see if the very robust findings of Value Added I still pertained and secondly, if growth on these dimensions could be found to be associated with particular collegiate experiences.

Since it is always difficult in the social sciences to determine cause and effect relationships, our aim was to provide indices of frequency. For example: if a student has particular strengths as measured by VAI measures and he or she participates in a designed series of activities, then there is likelihood that their development will increase in accordance with suggested probabilities. If we could find correlations, we could use them to encourage a student who is weak in Expository Writing to participate actively in writing programs, revise their papers according to faculty comments, do a great deal of writing during their four years. At a macro level, we wanted to be able to say to departments or even institutions: if you want to produce students who have achieved these objectives then there is a list of kinds of experiences which you should provide. Value Added II was a search for such relationships. This paper represents an initial attempt to report the findings on this part of our search. I am delighted to have an opportunity to present them to the Commission.

It has not been an easy task. We have searched through mountains of data for indices and activities associated with the measures of student growth. We've had our peaks and our valleys. On the day the Cliff Adelman called inviting me to address the Commission, I was on a high mountain top, having just found some exciting results. Let me sketch these findings

at the moment; they will be presented in a full fledged research report to FIPSE later this fall. Let me limit my remarks to four major dimensions--analytical abilities, learning of new material, moral development, and the ability to present ideas in written form. On your behalf, let me ask three questions.

Who learns the most? Students were divided into three categories on the basis of secondary school credentials: the lowest third on secondary school grades and College Board or ACT scores, a middle third and a highest third. The college experience helped all three groups gain on all dimensions, for the changes in Value Added II were very significant.[1] There were, however, differential outcomes. The lowest group gained most in fundamental skills writing and also in the ability to think effectively, which is so fundamental to the writing process. The highest third gained most on the dimension of analytical ability. The middle third typically fell in between.

How is learning associated with college activities? Those who learned the most were the Renaissance people; those who participated most extensively in activities during their college years. One caveat, however, should be mentioned: students can participate in somany activites that these distract from the major foci of the development. On some dimensions, growth was curvilinear rather than linear; that is, those with somewhat less than the greatest participation scored the highest. The group that was the second

---

[1]This is not to suggest that every student improved on each dimension. There is no evidence that simply growing older insures higher scores. Since there is no way to establish a control group, composed of those of equal talent and equal interest in things scholastic who do not attend college, our substitute has been to demonstrate that growth on these dimensions does not take place for everyone simply with aging.

most frequent gainer overall were those students who had strong academic interest. Many of our indices have a cognitive component so it is not surprising that grades are correlated with growth; these people gain the most from their college experience. In general, an interest in things intellectual was important in improvement on our measures.

The third group that gained the most were athletes. They started rather poorly on many of our measures, but their growth was quite substantial. Work at an outside job for pay was fourth. Being engaged in something career-related or having a job of particular relevance to that individual seemed to be helpful on some of our dimensions.

The poorest group in terms of growth along these dimensions were those who neither were engaged academically nor participated in any activites. For these people growth was essentially nonexistent. One does have to be engaged, it appears, in intellectual or extracurricular activities and better still, both, to grow.

Do colleges differ in their ability to produce changes? Yes, markedly. It depends on which of the dimensions we are talking about as to the effectiveness of which colleges are the greatest. Boston State, as you will note in Figure I, was the largest contributor to an improvement in grammar. Its students started very poorly and rose to very acceptable standards in four years.[1] On other dimensions such as analytical ability, schools such as Harvard had the most pronounced effect. In my completed report, we will present these changes in greater detail.

---

[1]Students, it should be added, were matched on other independent variables such as entrance scores and secondary school grades. The matching was within schools, not between.

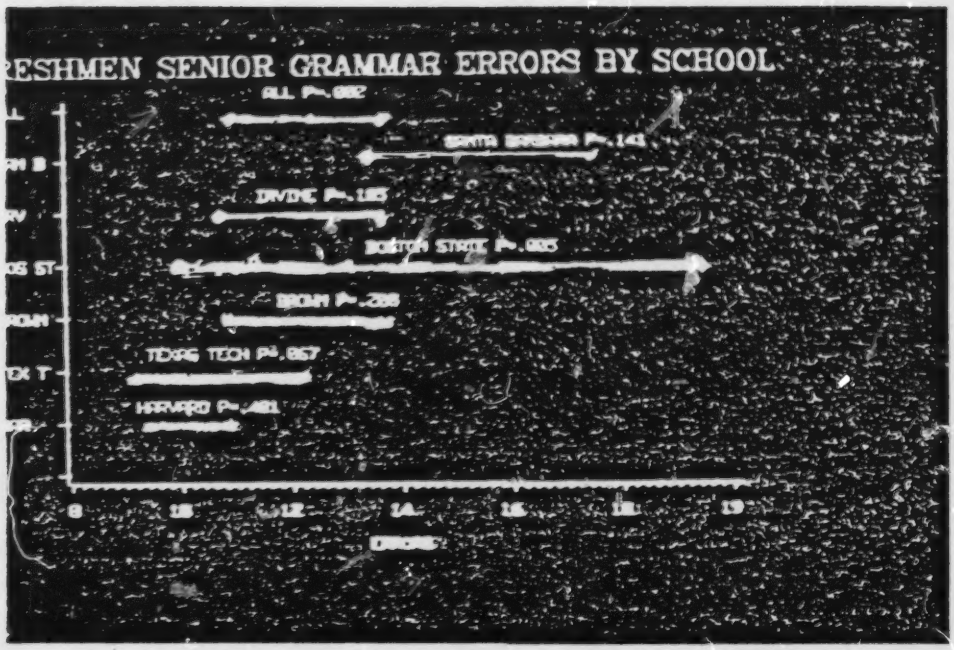


Figure I

Figure I represents the findings from the Logical and Rhetoric Test, it was scored on five dimensions--1) effectiveness of the paragraph as a prose statement, 2) the logic with which the arguments are developed, 3) the number of errors in grammar, 4) the number of incomplete sentences found in the passage, and 5) the number of misspelled words. The data in Figure I represents the third scale "grammar" which is simply the number of grammatical errors the students made in the three or four paragraphs they wrote in response to one of the sugeested topics. Like the scoring for a cross-country track meet, a small number of errors is to be preferred. Therefore the left of the graph is good; the right is bad. There is another way in which the scale is reversed: freshmen are on the right end of the red bar; seniors seniors are on the left. Seniors actually do write better than freshpeople and that includes their use of grammar. Starting at the top of the graph and working down, we can see that "all" seniors in our sample improved signi-gificantly over those of freshmen in their use of grammar (P .002). Univer-

sity of California-Irvine and Santa Barbara had significant improvement. Texas Tech did very well. The most remarkable change was that found at Boston State, an open admission school. Freshmen began very poorly compared to the rest of the sample; however, they ended very strong and competitive with any of the schools, hence the long length of their line. The length of the line represents the amount of change between freshmen and senior scores or the value added. Contrast that, if you will, with Harvard's red line which, while in the right direction, was relatively short.

What we have found in general in Value Added II is that colleges can and do add value in very large and significant ways. Although value added does depend upon the background of students, what they themselves did during the college years and which schools they attended, virtually all types of students profit from the experience along these measures.

Using the Value-Added methodology has given us insights into the learning experience that was unavailable by other approaches. Additional work still needs to be done. The measures we used are not necessarily perfect ones or necessarily right for all institutions. They demand a lot of student time to administer. They are cumber some to score. Scores aren't as reliable as one would like. But in spite of these limitations, the study has produced the strongest documentation we have of the effect of college on students.

Most of the colleges in the sample are continuing their explorations on their own using this approach. A number of other colleges have joined in pursuing these aims and explorations each using the portions of the materials that seem most appropriate to them. Several are now collecting their data by incorporating Value Added measures into the final exams in large courses. Potential for impact is high. I think that it would be enhanced if the

Commission were to recommend such an approach in their report.

One feature that definitely needs exploration is whether the changes produced by colleges are at the moment increasing or whether they are decreasing. That information can be acquired only by repeatedly measuring over time. The data currently available fails to address this very important point.

Let me now draw upon my experiences in admissions. I have been an observer of the changing secondary school scene for years, having served on the Harvard Admissions Committee for some 25 years and been the director of a Summer Institute on Admission, sponsored by Harvard and the College Board for almost as long. In the post-Sputnik era there seemed to be a virtually endless number of talented students. That bright situation has changed and I would like to address this decline in the next portion of my remarks.

The figures that follow use College Board data; the same findings would be true, however, if American College Testing or the National Assessment Program data were used.

In Figure II, you can see the marked drop in the Verbal mean scores over this period from 1952 to 1981. The high came in 1963 and they have been dropping for almost 20 years at an alarming rate. Yesterday, The New York Times ran a headline because the mean Board scores increased by a point. After two decades of decline even an increase of a point or two can be exciting. But like the stock market, is this the beginning of a bull market or simply a bear trap? The low scores are in my view a reflection of the poor quality of American education, just the reason that this Commission was established. The College Board itself established a committee headed by Willard Wertz to explore the causes underlying decreasing scores. They

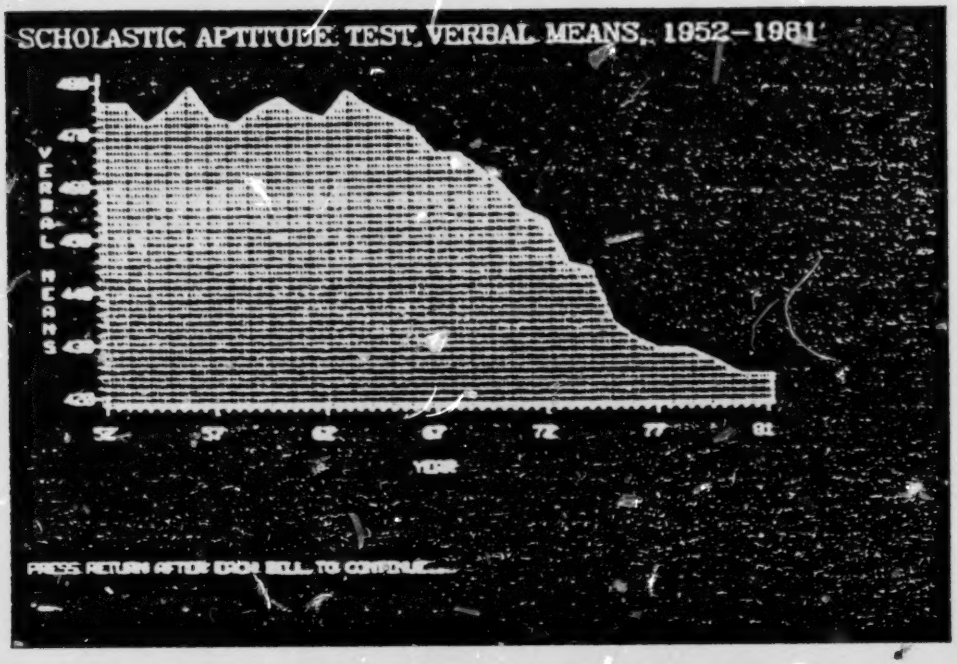


Figure II

reported multiple causes: the drop reflects in part the changing student populations taking the tests, the decreased time students spend reading (because of increased time in front of the television), the fact that textbooks used in secondary school are less demanding now than at an earlier period, and finally that motivation for things academic is lower than before. Let me add an aside about reading since we are looking at verbal scores, which have a high reading component. Last fall I visited LaFolette High School in Madison, Wisconsin. When I walked in the door it seemd very, very quiet for a modern high school. I asked if there were a holiday. In fact, everyone was reading. The school established a program where the whole school stopped all activities for 20 minutes each day and everyone read. My reactions were mixed; first I found it appealing because it was so direct and simple, but on the other hand, it seemed so trivial. Could 20 minutes of reading possibly have any influence? Upon further reflection, I decided that

there are probably many students who do little reading and that this may represent an increase of time of 20%, 40%, maybe even 60%. Not so paradoxically, the reading scores at LaFollette have been going up since this program was established.

Some of the papers prepared for the Commission have given the impression that it is the poor student in poor schools who is the source of poor achievement in American secondary education. The evidence from my admissions experience led me to quite a different opinion. Our failures in schooling have been across the board. In Figure III you see the number of students nationally who, between 1976 and 1981, had SAT scores Verbals in the 750 to 800 score range. These years are not ones of a typical decline in the

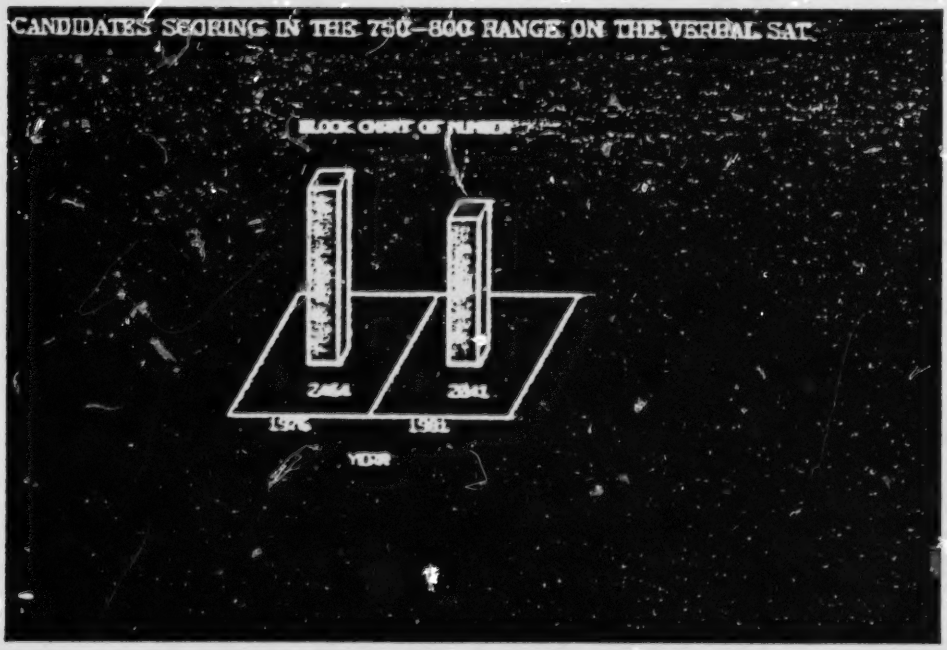


Figure III

in the SAT, yet the graph represents a five year span during which there was a drop of approximately 400 people; or about 20% of the population taking the

test nationwide.

Figure IV represents the decline in SAT-M mathematics scores over the period of 1952 to 1981 and you can see that it has been pronounced as that for the SAT-V.

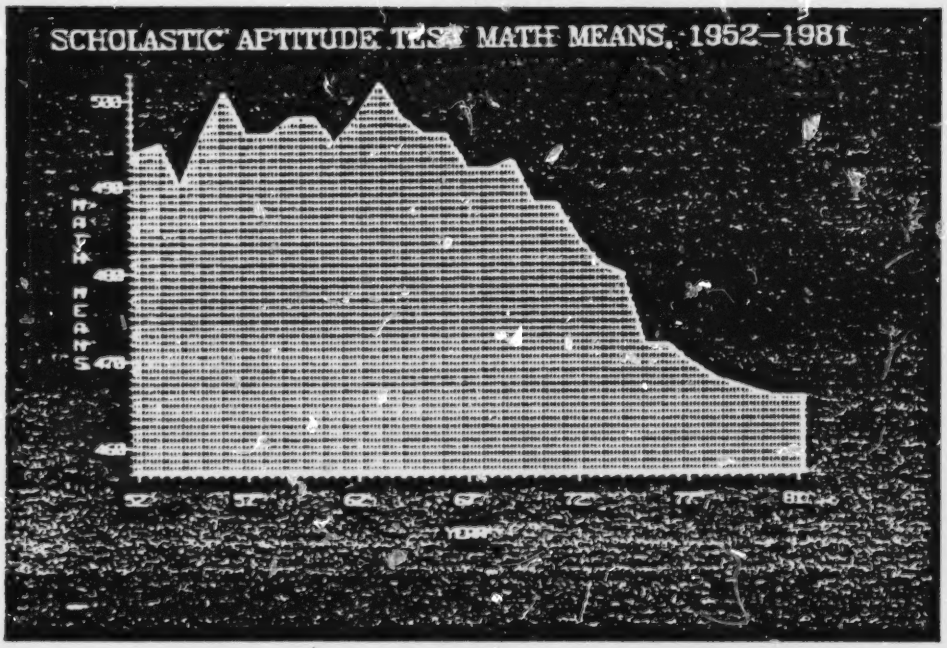


Figure IV

We compared the group scoring 750 to 800 on the SAT-M in 1976 and in 1981 (unfortunately no figure is available). There was an enormous loss of talent in just this five year period; the number dropped from approximately 12,000 students to a little over 6500.

These changes should be juxtaposed against the new research that has been done on Japanese students by Richard Lynn, a British psychologist: he found that they have been having gains that are comparable to our losses. He translated a commonly used American intelligence test into Japanese and found that the Japanese children had an average IQ of 111 versus that of 100 for

Americans. Further, he found that 10% of the Japanese population had IQ's over 130 as compared to 2% in America and Western European countries. This work is controversial, but the controversy has centered not around the fact that the Japanese average IQ is 111 or that 10% of their population have scores of 130 or more. The controversy has been about the causes of these discrepancies. Some have argued they are genetically driven, others have argued that it is a function of education. For our national interests and for the interests of our students, we must hope that it is the result of schooling.

Let's look at Achievement test results. For a number of years, the Achievement tests have been a better predictor of grades at Harvard than have the SATs. The regression equations computed this spring were particularly interesting: the results showed that the scores on Achievement tests were not only superior to the Scholastic Aptitude tests, but also were more important as predictors than secondary school grades. This is the first time that has ever been true; further the total power found in their predictive equation was also higher than it has been for almost two decades. I've always been rather pleased to be able to claim that it helped to "know something" to get good grades at Harvard; this evidence shows that it is especially true now. I think from a national standpoint that Achievemnt measures are a very important commodity in the educational market place and it is worth taking a serious look at the results of these tests.

In the following figures, the most recent achievement scores falling in the 750 to 800 range are juxtaposed against those of 5 years ago. First, let's examine a series of foreign language scores. In Figure V you can see that few scored in the 750-800 score range in German in 1976 and that small number dropped markedly in 1981. If we look at French (Figure VI), a very

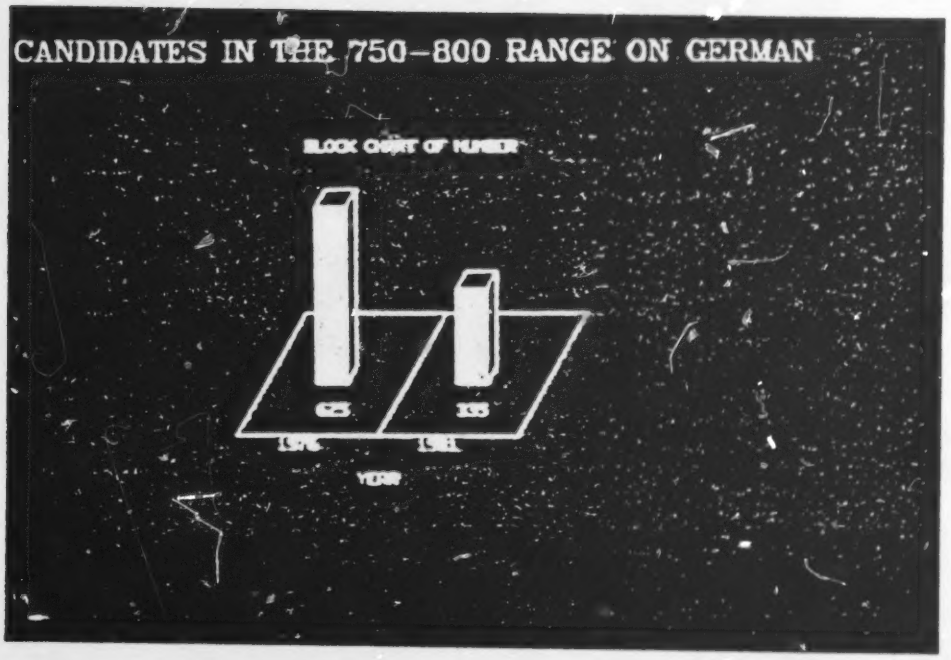


Figure V

popular language in secondary school (if there are popular foreign languages anymore), there is an equally precipitous drop. In Figure VII, which gives

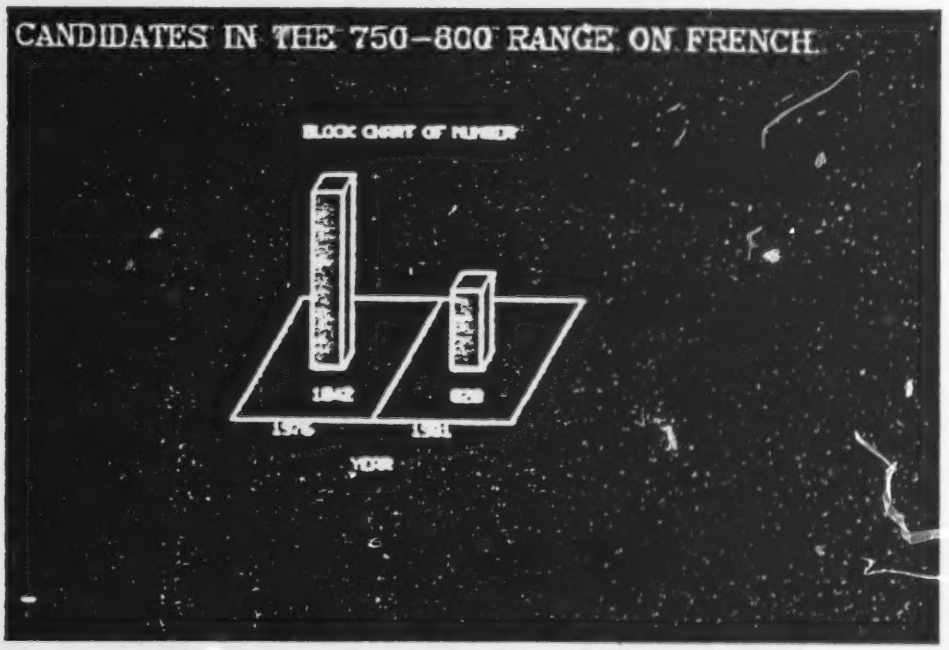


Figure VI

the results of the Spanish test (Spanish students are surely the nation's fastest growing group of language students), we again see this staggering drop at the upper score ranges.

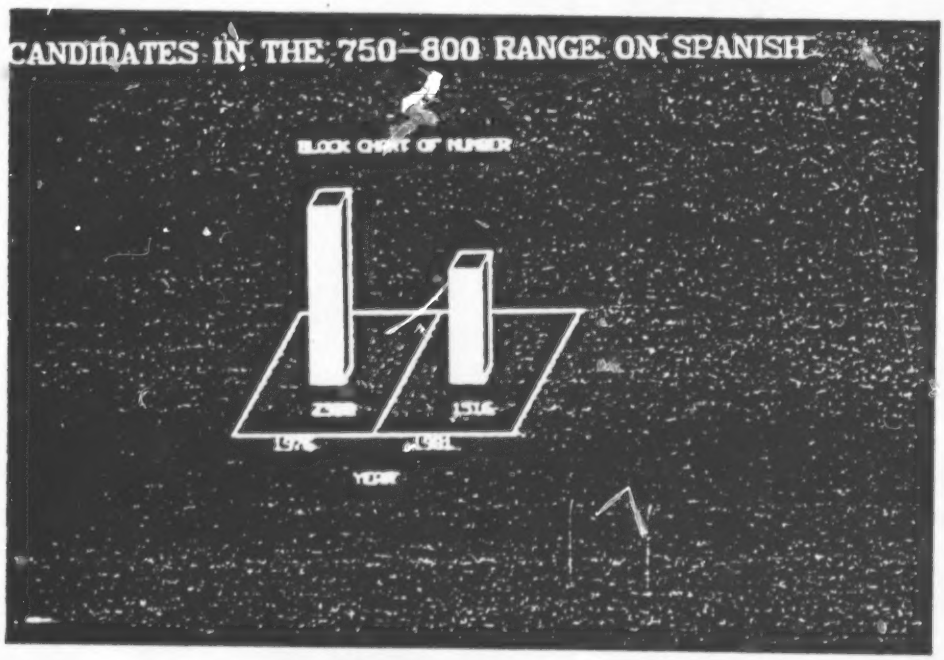


Figure VII

Results from our own language, English, are found in Figure VIII. This repre-

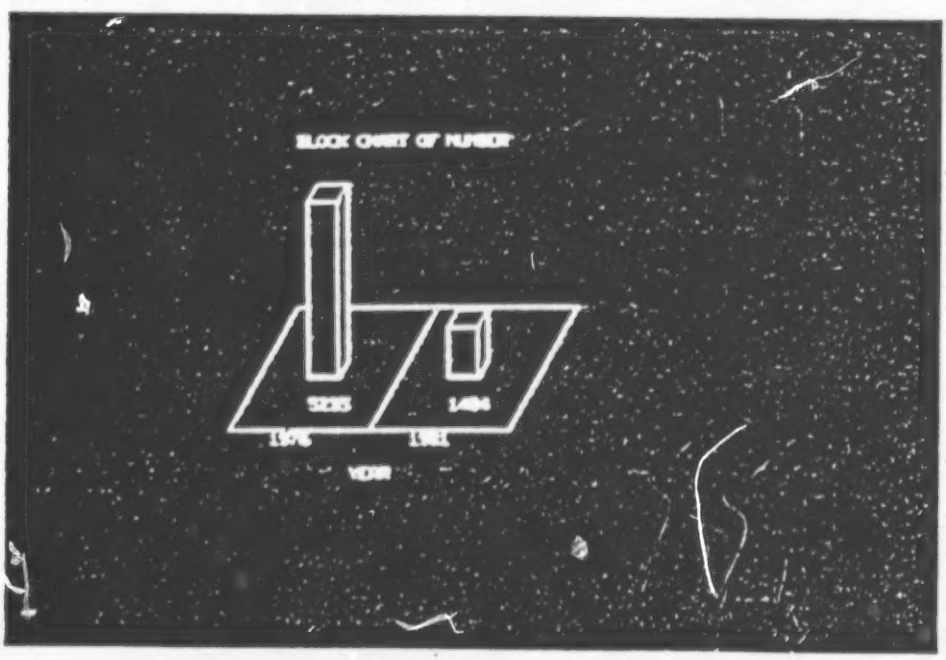


Figure VIII
CANDIDATES IN THE 750–800 RANGE ON ENGLISH COMPOSITION

sents an appalling drop in student ability to handle language at a sophisticated level. Reflecting back on the Value Added material presented in Figure I we have demonstrated that it is possible to teach grammar and other components of language in college. We can do it for students who are poorly prepared, but is that an appropriate role for colleges?

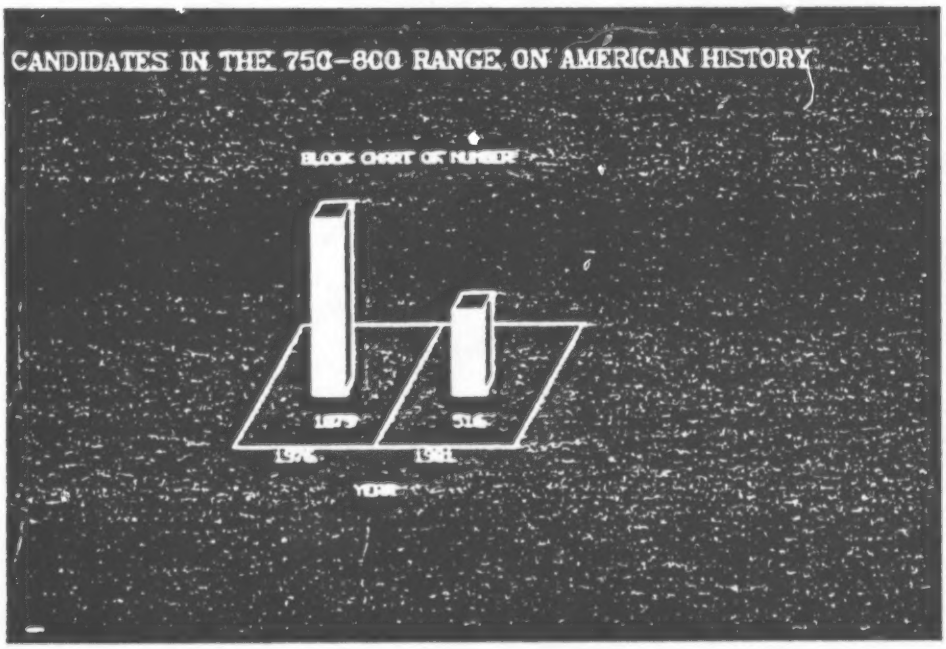
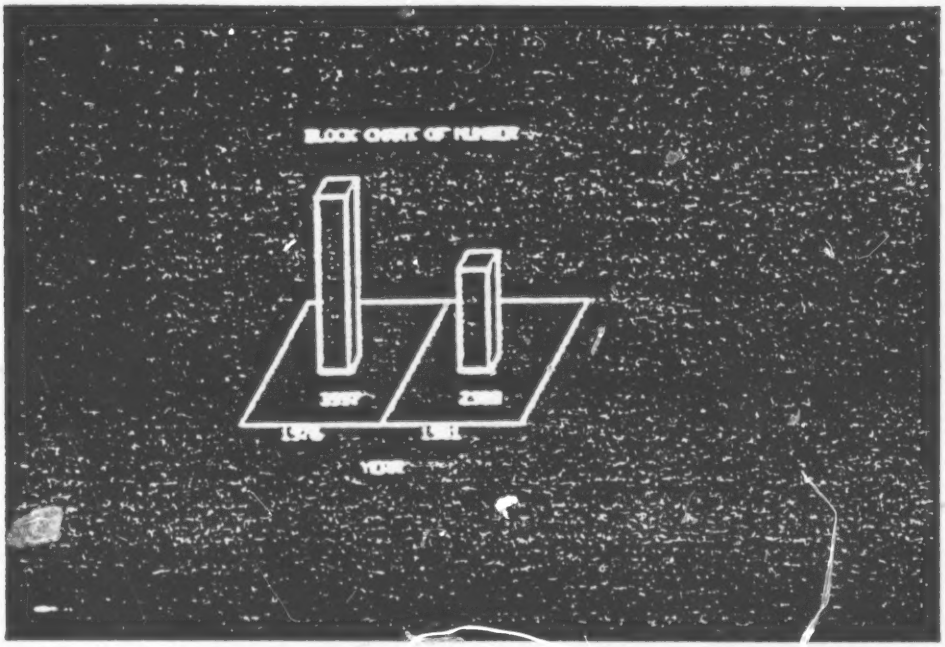


Figure IX

Examining more of the Achievement series in Figure IX, we can see the sizeable drop that took place in American History. In Math Level I, Figure X demonstrates an equally appalling drop during just this five year period. Biology, Figure XI, has also dropped alarmingly; five years ago there were over 3000 who scored over 750 at that time. I'm delighted to say that (Figure XII) there has not been a corresponding drop in Physics. Unfortunately, if we were to go back to an even earlier period, we would find that there had been a drop in Physics, too.



CANDIDATES IN THE 750-800 RANGE ON MATH LEVEL I

Figure X

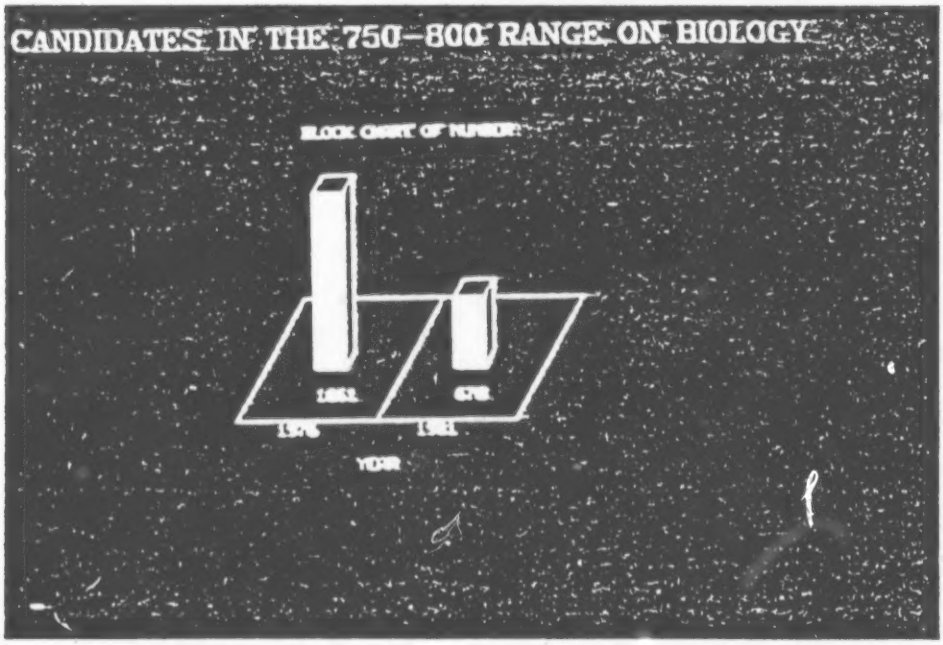


Figure XI

CANDIDATES IN THE 750-800 RANGE ON PHYSICS

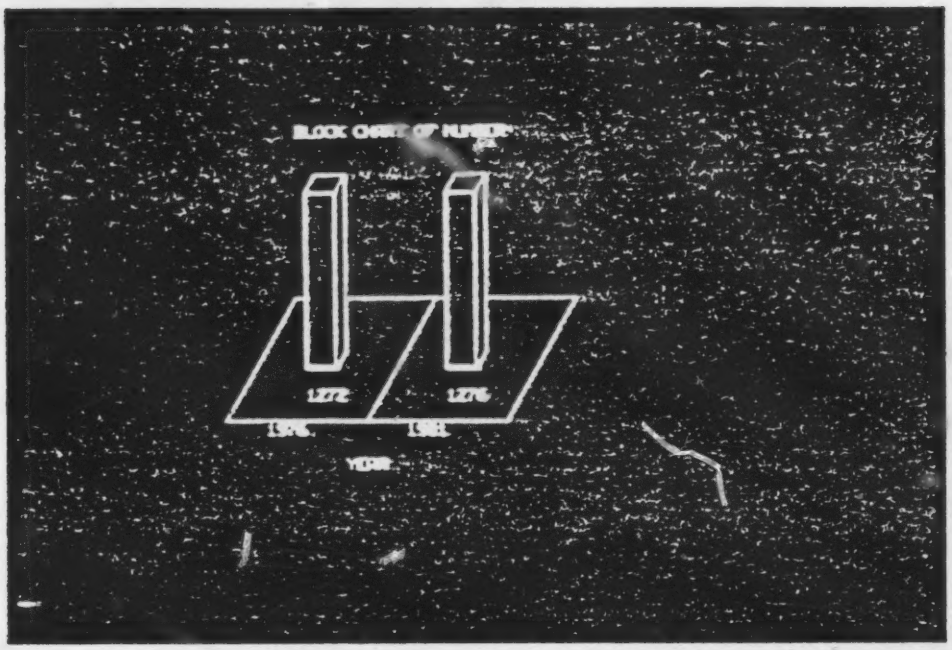
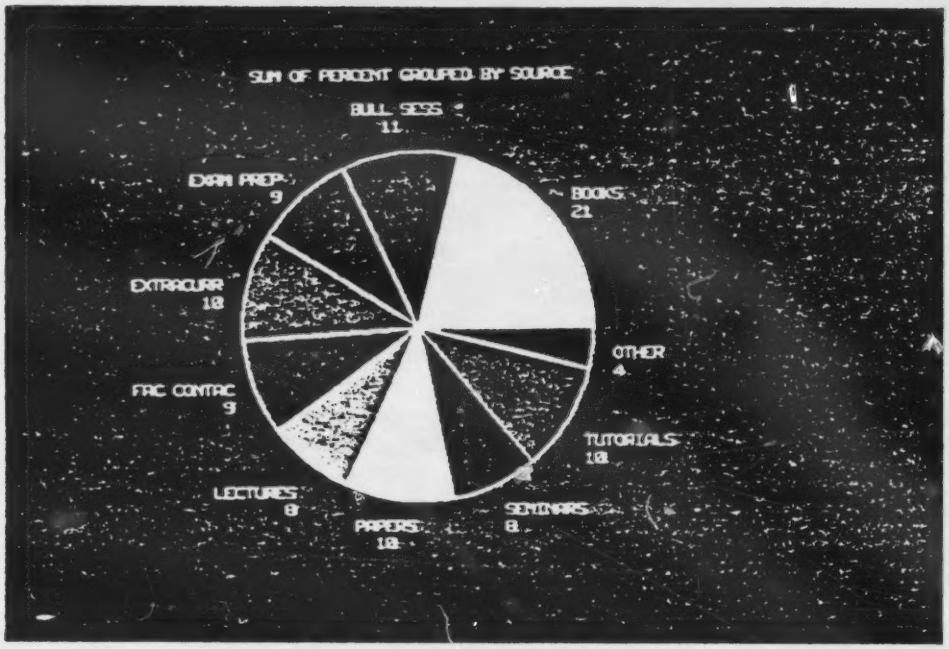


Figure XII

Having long before this become concerned about these dramatic changes--that they were taking place across the board, that they involved not simply poor students in poor schools, but for all kinds of students in all schools--I wanted to turn the discussion to the sources of learning. How do students acquire or fail to acquire the information, the understanding, the judgement and the analytical powers we are talking about? I asked them to indicate the forms that were most effective in contributing to their learning. That information you see in Figure XIII. These data on students' opinions was collected in 1981; I also collected the same type of data in 1971, a decade earlier, and none of the percentages differed by more than 1%. How students learn hasn't changed. The figure becomes even more interesting if you add some of the pie shaped wedges together. For example, if you add together books and papers, which I would call "private learning", you



SOURCES OF LEARNING AS ATTRIBUTED BY STUDENTS

Figure XIII

can see the overwhelming predominance of this instructional mode. In a sense, one could toss lectures into this category because they constitute a motivating force for student reading and writing. The second most prominent style of learning occurs through several types of interpersonal activities: seminars, tutorials and bull sessions. Interactive, interpersonal opportunities rank second only to the private learning in effectiveness as a learning mode. What this suggests is that today just as a decade ago, students need encouragement and opportunity to do their own private, directed work, and secondly, to have opportunities of interpersonal interchanges to supplement and reinforce their private work.

Figure XIV is limited to Harvard graduates, but I think it is equally applicable nationally. Along the bottom of the graph are dates when the information was collected, namely, from the classes of '67, '71, '74, '77 and '81. On the left hand axis are the vocations they chose. The first

VOCATION OF GRADUATES

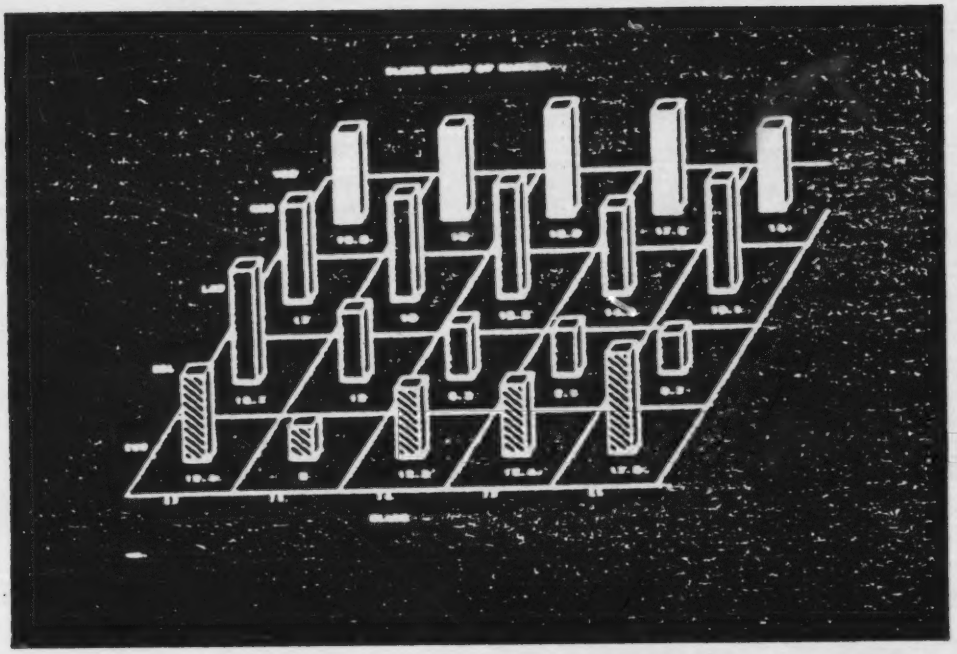
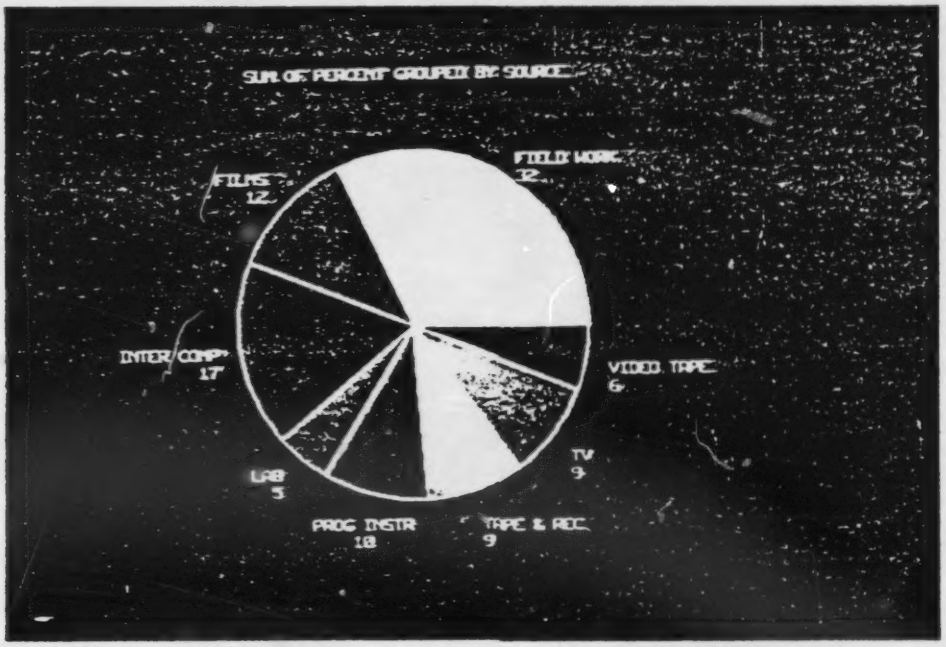


Figure XIV

series of columns in crosshatched red represent business, the second column (green) represents college or secondary school teaching, the third (yellow) law, and the fourth (pink), medicine. During this span of a bit more than a decade, we can see that business interests of graduates dropped and then grown rather markedly and there has been a sharp and continuous drop in careers in education. Law and medicine have remained essentially constant at a rather substantial levels. All of us are aware and concerned about the fact that the number of students choosing to follow academic careers in higher education or teach in secondary schools is appallingly low. If we reflect on the fact that learning takes place through interaction with great teachers, hopeuflly at the seminar or tutorial level, if not at least in a large lecture, we must conclude that resources are just not going to be available for students in the near future. Some research recently complete at the University of Washington showed that not only are those enrolling

in a program heading to a career in education the poorest students academically, but that those who complete their degrees in education are the poorest of students who originally enrolled in education.

If we were to accept the fact that there are going to be limited human resources of outstanding quality in the teaching profession (something we must change), are there other ways that we can begin to conduct our educational system so as not to shortchange students of the next generation? Searching for alternatives, I asked college graduates how effective they thought the various types of media were as instructional tools. The results are found in Figure XV. I asked them to check any media with which they had



EFFECTIVENESS OF MEDIA AS AN INSTRUCTIONAL MATERIAL

Figure. XV

had experience and the degree to which it had been effective. The experience could only have been in either secondary school or in college. We see that the pattern is far from clear. Films are helpful, but not for

very many. TV is helpful, but not for very many. Programmed instruction which a decade ago was thought to be the answer to our instructional problems, clearly has fallen short of its mark. As of yet the interactive computer has not turned the tide. It might have potential, but the software in the experiences of these students has not lived up to the exciting promises.

The productivity of the schools is dropping; it takes good teachers and motivated students to change the current direction; the human resources are not being prepared and technology has provided no substitute. I hate to end on such a dismal note. Let me try to sum up my reactions and make some suggestions to the Commission.

First, there are reasons for serious concern about American education. It is failing us across the board--all types of schools and all kinds of students. Harold Howe, the former U.S. Commissioner, feels that the youth culture was never more pronounced or more separated from the adult culture than it is today. If we were simply to make the curriculum more rigorous then we would simply bypass even more students. He feels that we should change the style of instruction to get more students into the mainstream. Ken Clark, in a recent New Yorker profile, reiterated his stand, that we judge teachers by the value they add to their students' knowledge, and demand that all children know and use standard English and not that of the street.

Personally, I wish that the Commission could find a way to regain some of the excitement that we found in the schools during their period of curriculum reform. Many teachers in secondary schools and colleges need retraining. We need science in the elementary school, we need a new emphasis on reading and writing, figuring and languages. It will take an enor-

mous infusion of federal leadership to bring these things about. We need a new surge of national enthusiasm for these goals.

Second, Value Added is a useful tech ique. It is a strong addition to the traditional evaluation approaches that are widely used. Feedback is necessary for survival. We know that if a person is placed in a sensory-deprivation vat; so that he or she cannot feel, touch, smell or have any feedback from the senses, in less than an hour he or she begins to exhibit psychotic symptoms. Institutions have an equally pronounced need for feedback. Value Added is a substantial improvement in the way such feedback can be obtained. While I've introduced the Value Added methodology to a number of colleges, that is far from enough. It is equally important at the secondary school level. It is an enormously effective mechanism for change. Unfortunately, it is not as simple as Alexander Astin suggested; using teacher-made tests as pre and post measures is not adequate. If a teacher wanted to insure that great change resulted from his or her teaching, he or she would simply create tests where the questions seek information taught solely in that particular class. Unfortunately, this can be most easily accomplished by testing for simple information. We often times hear the criticism now that teaching is geared to the tests. In this case, the tests would be geared to the teaching. Good tests and good teaching go hand in hand. We must insure that this is the path that takes place.

Let me go a step further; I believe there is a role for a federal scholarship program on merit. Let the states worry about minimal competency examinations to insure that students are entitled to a high school diploma. Let the federal governement reward the best and the brightest. If there were a series of national examinations designed to stretch and reward the

most talented of our students, it might rekindle the pursuit of excellence in the schools.

My third suggestion was addressed in the panel discussion on August 27th. I recommend that you review and integrate the findings of the studies currently being conducted on American education with the aim of defining the intellectual tasks of schools and colleges. We should not ignore the developmental stages of growth, but it must be balanced with the cognitive development. Much work has already been done. At Harvard a Core Curriculum program has recently been created. The rationale for this program is now recorded in an excellent book by Phyllis Keller to be released this fall. Jerry Gaff, in his General Education Models provides another excellent source. Zelda Gamson is currently writing a definitive piece for the Commission that will explicate what higher education should be in this portion of the 20th century. I'm sure that you have been reviewing the 10 or 12 studies currently in progress on secondary schools. To integrate these studies on secondary and higher education would be of enormous service to your constituency, the American public. To give these statements the imprimatur of a national commission would be very important. Individual institutions feel helpless in making recommendations. Yale just reinstituted language requirement. To how many schools does that make a difference? When Harvard put in its Core Curriculum requirement, it also added a quantitative reasoning requirement and a computer requirement. That, too will have a positive but relatively small influence. Phillips Academy Andover will be aware of the change that Harvard made, but only a miniscule faction of the 25,000 other secondary schools of the country will be aware of it. A similar recommendation made by the Commission could have an enormous influence.

The Commission should recommend the finest in American education while avoiding the straight-jacket of a national curriculum. As it has been suggested, guidelines should exist, but they should not be so tight that one can't see the light through the crack in the door.